Emerging role of microRNAs in regulating long non coding RNAs in Breast cancer

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE
IN
LIFE SCIENCE

SUBMITTED TO
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA

BY
KALPANA DAS
ROLL NO. 412LS2035

UNDER THE SUPERVISION OF
DR. BIBEKANAND MALLICK
ASSISTANT PROFESSOR

DEPARTMENT OF LIFE SCIENCE
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769 008, ODISHA, INDIA
2014

राष्ट्रीय प्रौद्योगिकी संस्थान, राउरकेला
National Institute of Technology, Rourkela

Dr. Bibekanand Mallick, M.Tech., Ph.D.
Assistant Professor

RNAi & Functional Genomics Lab.
Department of Life Science
National Institute of Technology
Rourkela - 769 008, Odisha, India
Telephone: +91-661-246 2685 (O)
E-mails: vivek.iitian@gmail.com,
mallickb@nitrkl.ac.in
Website: http://vvekslab.in

CERTIFICATE

This is to certify that the thesis entitled "*Emerging role of microRNAs in regulating long non-coding RNAs in breast cancer*" submitted by **Kalpana Das** (Roll No: 412LS2035) in partial fulfilment of the requirements for the award of Master of Science in Life Science to the National Institute of Technology, Rourkela is an authentic and original record of research work carried out by her under my supervision and guidance.

To the best of my knowledge, the work incorporated in this thesis has not been submitted elsewhere for the award of any degree.

10/05/2014

Place: Rourkela

Date: 10/05/2014

(Dr. Bibekanand Mallick)

Dr. Bibekanand Mallick
Assistant Professor
Department of Life Science
NATIONAL INSTITUTE OF TECHNOLOGY
Rourkela-769008, Odisha

ACKNOWLEDGEMENTS

A project is a golden opportunity for learning and self development. I consider myself very lucky and honoured to have so many wonderful people lead me through in completion of this project. I take this opportunity to thank my Almighty God for His abundant grace which enabled me to finish my project work successfully.

I express my heartfelt gratitude & indebtness to my Guide, **Dr. Bibekanand Mallick,** assistant professor, Dept. Of Life Science, National Institute of Technology, Rourkela for his valuable guidance, encouragement, patience, & worthy suggestions for my project work, without which this project work would not have materialized.

I also convey my deep gratitude to Prof. S.K. Sarangi, Director, National Institute of Technology, Rourkela for providing excellent facilities in the Institute for carrying out research.

I would like to take the opportunity to acknowledge quite explicitly with gratitude my debt to all the Professors and Staffs, Department of Life Science, National Institute of Technology, Rourkela for their generous support throughout my project work.

I am grateful to PhD scholar Mrs. Devyani Samantarrai, Ms. Mausumi Sahu, Ms. Debashree Das, Mr. Bedanta Ballav Mohanty, and Mrs. Garima Singh for their support & guidance during my project work.

I am indebted to my loving parents, my near & dear ones for their inspiration, support & love. Finally, special thanks to all my friends & well-wishers for their support throughout my course & project work.

(KALPANA DAS)

LIST OF TABLES

LIST OF FIGURES

CONTENTS

ABSTRACT

Genome wide analysis reveals that only 2% of the genome is translated which gives the functional protein product. The remaining 98% do not code for any proteins and transcribed as non-coding RNAs (ncRNAs). These ncRNAs are involved in diverse biological processes. When these are deregulated, they cause cancer. microRNA (miRNA), a small ncRNA has been reported to be involved in cancer, but recently long non-coding RNAs (lncRNAs) are also emerged to be playing a role in cancers. LncRNAs are dysregulated in various types of cancer which includes breast, gastric, hepatocellular carcinoma etc. In this work, we considered an lncRNA, HULC (highly upregulated in liver cancer) to study its expression in breast cancer (which was not reported earlier) followed by studying its interaction with a miRNA, hsa-miR-30a-3p. From our qRT-PCR analysis, it is found that HULC is up-regulated and miRNA is down-regulated in breast cancer cell line, MDA-MB-231. Observing this inverse correlation in their expression followed by finding miRNA target sites within HULC, it was hypothesized that down-regulation of hsa-miR-30a-3p might be leading to the up-regulation of HULC in breast cancer. This reports a case where one small ncRNA regulate an lncRNA which might have implications in breast cancer.

Key Words: microRNA, lncRNA, RNA-seq, microarray, breast cancer

INTRODUCTION

A hundred years ago only few people were died of cancer. However, now cancer has uncountable bad impacts on our health. This has unfortunately become a familiar occurrence and is counted as one of the ten leading causes of death in India. This might be due to day by day exponential rise in environmental toxins from foods, water and air which has sadly become a part of our life. Cancer is a class of diseases characterized by uncontrolled cell growth forming malignant tumors, and invading nearby parts of the body. However not all tumors are cancerous; benign tumor do not invade normal tissues and do not spread throughout the body. Among several types of cancer, breast cancer is second leading cause of mortality (Babu et al. 2013).

Worldwide, breast cancer is now the most frequently diagnosed cancer among women (Redig et al. 2013) both in the developed and developing countries. Although breast cancer is thought to be a disease of the developed countries, almost 50% of breast cancer cases and 58% of deaths occur in less developed countries also. 519,000 deaths were occurred due to breast cancer in breast cancer, worldwide (7% of cancer deaths and almost 1% of all deaths). It is estimated that 1.38 million new cases were diagnosed in 2008 (23% of all cancers) and around 458,000 deaths have occurred due to breast cancer (Ferlay et al. 2010). In 2014, it is estimated that, there will be 232,670 new cases of invasive breast cancer , 62,570 new cases of in situ breast cancer which includes ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS). Of those, about 83 percent will be DCIS. 40,000 breast cancer deaths will be occurred (Adopted from American Cancer Society**).** So breast cancer has a devastating effect on our health.

There are several causes of breast cancer. Now-a-days, non coding RNAs (ncRNAs) play a major role in development and progression of breast cancer due to their regulatory functions in transcriptional, post-transcriptional and epigenetic events. NcRNAs are deregulated in number of cancer types such as breast, lung and colorectal cancer etc demonstrating their oncogenic role in cancer systems and has been linked in cell proliferation and migration. These ncRNAs include small ncRNAs (up to 200 nt in length) as well as long non coding RNAs (from 200 nt up to 100kb) **(**Gibb et al. 2011).

Small non coding RNA includes microRNA, piRNA, and siRNA. All sncRNA when deregulated cause cancer. MicroRNAs have emerging role in cancer system. MicroRNAs (miRNAs) regulate gene expression post-transcriptionally by binding to the 3'untranslated (3'UTR) region of target messenger RNA (mRNAs) and inhibit the translation or induce

degradation of mRNAs (Bartel 2009). MicroRNAs such as miR-10b, miR-125b, miR-145, miR-21, miR-155, and miR-181 were revealed to be the most consistently deregulated miRNA in breast cancer.

In addition to small ncRNAs like microRNAs, long non coding RNAs (lncRNAs) has been is emerged out recently to have potential role in cancer development. LncRNAs play immeasurable role in several biological processes (Martin etal, 2014) which include X chromosome inactivation, nuclear structure, genomic imprinting and development (Fenoglio et al 2013). Dysfunction of lncRNAs has been strongly associated with several human diseases mainly cancer which includes modulation of melanoma cell apoptosis and invasion by SPRY-IT1 (Khaitan et al. 2011), promotion of bladder cancer proliferate by UCA1a, control of breast cancer cell apoptosis by GAS5.Deregulation of lncRNA may happen due to miRNA or some other agents. There is less report about lncRNA and miRNA interaction that includes MALAT1, which is down regulated by miR-9 (Leucci et al. 2013), HULC, which is up regulated due to miR-372 (Wang et al. 2010).

In our present study, we endeavoured to find novel miRNA-lncRNA interactions which has crucial role in breast cancer. We did real time PCR for both miRNA and lncRNA for expression study using breast cancer cell line MDA-MB-231 and normal cell line HaCaT. We got lncRNA HULC from RNA seq data (which is up regulated) and miR-30a-3p from RNA hybrid. HULC is upregulated in liver cancer as name suggests but later on it is also over expressed in gastric cancer. We want to find out whether HULC is over expressed in breast cancer or not. So we did expression study of HULC through wet lab to see whether it shows the same result to that of RNA seq data. Then we did interaction study whether one non coding RNA regulates the other. In this study, in the absence of miRNA, lncRNA is up regulated and in its presence it is down-regulated. So regulation of lncRNA is dependent on miRNA and its down regulation due to miRNA pays a pivotal role in breast cancer.

REVIEW OF LITERATURE:

Normally, when there is a requirement of more cells to the body, older cells get died and younger cell divide to form new cells. But sometimes when cells are not needed, cells also divide and this extra mass is called tumor. These tumors are of two types i.e. benign tumor and malignant tumor. Benign tumors do not invade the body parts where is malignant tumors invade the body parts and is more dangerous. This malignant tumor is called cancer.

Cancer is a malignant tumor involving uncontrolled cell growth. More than 100 types of cancers are seen which includes breast cancer, skin cancer, lung cancer, colon cancer, prostate cancer, and lymphoma etc. Symptoms of different types of cancer vary depending upon the cancer type. Chemotherapy, radiation or surgery is some of the methods of treating cancer. Cancer is characterized by several key characteristic features which are known as hallmarks of cancer proposed by Hanahan and Weinberg. These are described below in figure 1.

Hall marks of cancer:.

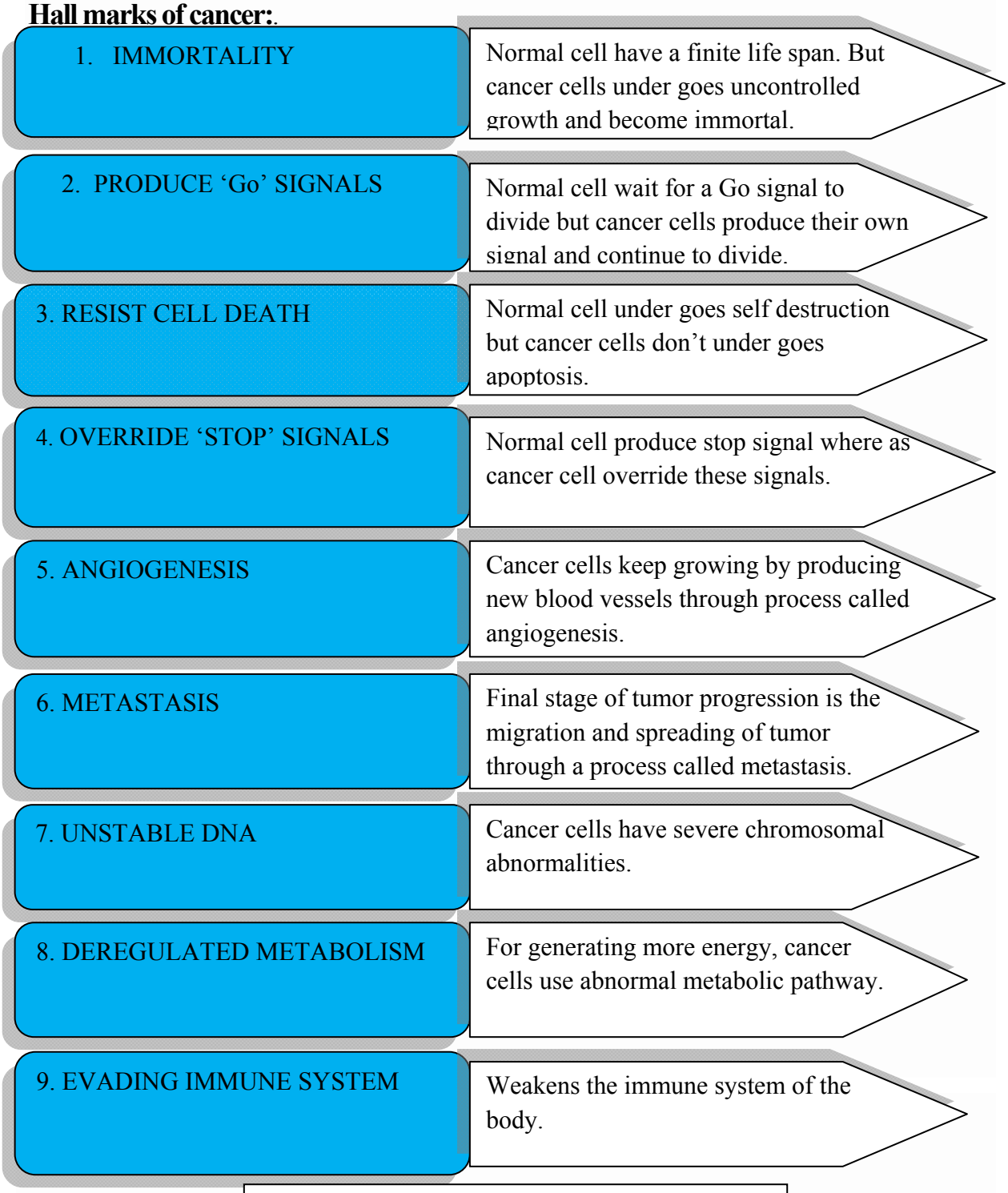

Fig. (1). Hallmarks of cancer

Breast cancer is a type of cancer originating from the inner lining of milk ducts or the lobules that supply the ducts with milk. Cancers originating from ducts are called as ductal carcinomas and that from lobules are called as lobular carcinomas. Breast cancer can occur in human as well as other mammals. Majority of breast cancers are seen in female and rarely

in male. It is more than 100 times more common in women than in men Breast cancer accounts for 23% of all cancers in women. 458,503 deaths were occurred in 2008 which accounts for 13.7% cancer death in women. But survival rates in the Western world are high that is more than 8 out of 10 women (85%) in England diagnosed with breast cancer survive for at least 5 years. Survival rate in developing countries are poor due to poor prognosis.

Breast cancer can be invasive or non-invasive. In invasive cancer breast cancer, the cancer spreads outside the membrane of the lobule or duct into the breast tissue. It also can then spread into the lymph nodes in the armpit or beyond. Non-invasive breast cancer is the cancer which is still inside its place of origin and has not broken out. Invasive ductal carcinoma is the most common type of breast cancer

Symptoms of breast cancer:

According to the National Health Service, UK, women who detect any of the following signs or symptoms should tell their doctor:

- Lump formation in breast.
- Pain in the armpits or breast.
- Redness of the skin of the breast like the skin of an orange.
- Bleeding can occur from the tumor.
- Rash around the nipples.
- Swelling in one of the armpits.
- An area of thickened tissue in a breast.
- One of the nipples has a discharge; sometimes it may contain blood.
- The nipple becomes sunken or inverted.
- The size or the shape of the breast changes.
- The nipple-skin or breast-skin may have started to peel, scale or flake

Causes of breast cancer:

Body's breast cells also die and new cells come out like that of other cells. There is always wear and tear. When deviation occurs from these phenomena, breast cancer occurs. Followings are some of the causes of breast cancer.

- Women over 50 are more prone to breast cancer.
- If a woman's mother or sister had the disease before menopause, this is occasionally associated with one of two genes linked to breast cancer.
- Previous breast cancer.
- Women who haven't had children, or whose first child was born after age 30.
- Age of menstruation - starting periods at a young age (under 12 years old)
- Women who smoke regularly are also prone to breast cancer.
- Women with dense breast tissue.
- Radiation treatment to the chest before the age of 30.
- More alcohol consumption contributes to the risk of breast cancer.
- Uses of oral contraceptives increase risks slightly.
- Obese women who take excess caloric and fatty food.
- Two genes BRCA1 and BRCA2 responsible for some instances of familial breast cancer.
- More exposure to hormone estrogen, more breast cancer cans occur. Estrogen gives signal to the cells to divide, so the cell becomes abnormal in nature.
- Deregulation of non coding genes.

Non-coding RNAs:

There are only ~20,000 protein coding genes which represents only <2% of the total genome (Stein. 2004). This fact represents that only 2% is translated and 98% is transcribed (Birney et al. 2007). These transcribed regions are called as non coding region. Non coding RNAs play a major role in developing cancer when they get deregulated. These ncRNAs include small non coding RNA (sncRNA) and long non coding RNA (lncRNA). SncRNA includes microRNA (miRNA), small interfering RNA (siRNA), piwi interacting RNA (piRNA). Among them miRNA is most thoroughly studied ncRNAs and has extensive role in cancer development. LncRNA includes promoter associated RNA, antisense RNA etc. LncRNA includes many functions like transcriptional and post-transcriptional regulation, subcellular

trafficking and organelle biogenesis. When ncRNAs get deregulated it causes many kinds of cancer.

Now-a-days non coding RNAs play a major role in development and progression of cancer due to their various functions in transcriptional, posttranscriptional and epigenetic mechanisms of gene regulation. NcRNAs has been linked in cell proliferation and migration in cancer system like breast; lung and colorectal cancer etc. ncRNAs are deregulated in number of cancer types demonstrating their oncogenic role in cancer systems. This aberrant expression of ncRNAs may contribute substantially to cancer development. This ncRNAs include small as well as long ncRNAs. Small non coding RNA like microRNAs has role in cancer development but recently long non coding RNA is emerged out.

MicroRNAs:

MicroRNAs (miRNAs) are an abundant class of small ncRNA govern the post-transcriptional gene silencing through translational repression, and decay of their target messenger RNAs (mRNAs) by binding to its 3'UTR (Guo H et al. 2014). They are single-stranded RNAs having length 18-24nt. (Bartel et al. 2014).They play critical roles in diverse biological processes in plants and as well as in animals such as embryogenesis, cell differentiation, cell proliferation, organogenesis, apoptosis, and tumorigenesis. The biogenesis of miRNA in plant and animal is different (fig-2). Intergenic miRNA genes are transcribed into primary miRNA (pri-miRNA) transcripts by either RNA pol.II or RNA pol.III. In animal pri-miRNA is cleaved into precursor miRNA (pre-miRNA) Drosha. This pre-miRNA is then exported into cytoplasm by exportin 5 and is cleaved by Dicer. The product which is formed by this cleaving is called miRNA-miRNA* duplex having length ~22-nt. One mature miRNA is then binds to Ago-2 protein to form the RISC complex which is used for mRNA degradation. In plants, instead of Drosha, Dicer like-1 (DCL-1) cleaves the pri-miRNA to miRNA-miRNA* duplex (Song L etal, 2007). One mature miRNA then binds to

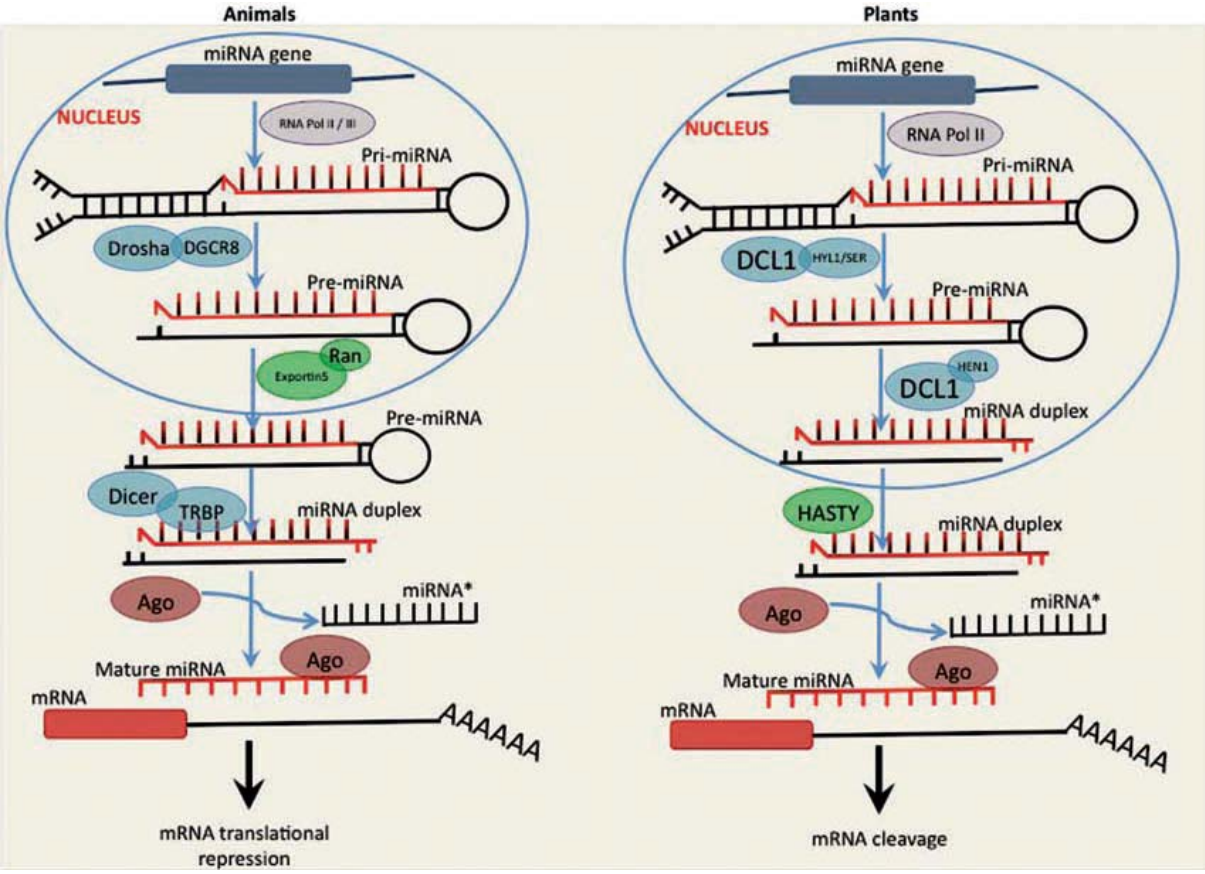

Fig. (2). Biogenesis of miRNA (Mallick et al. 2011)

miRNA targeting:

In miRNA, 2-7nt from the 5'end is called "seed" region. This seed region is complementary to the 3'UTR of the target. If this region is fully complementary to the target, complete decay will occur. MiRNA should be positioned within the 3′ UTR at least 15 nt away from the stop codon and away from the center of long UTRs. The target site should be AU rich. High complementarity is seen in Watson-Crick (WC) base pairing. Apart from WC base pairing wobble pairing is also seen but WC pairing is most efficient one. There are three types of target sites viz. Canonical site, marginal site, atypical site. Canonical site consists of 7mer-A1, 7mer-M8, 8mer; marginal site contains 6mer and offset 6mer; atypical site contains 3' supplementary site and 3' complementary site (). Efficacy of targets is in order of 8mer>7mer-M8>7mer-A1>6mer>offset 6mer.

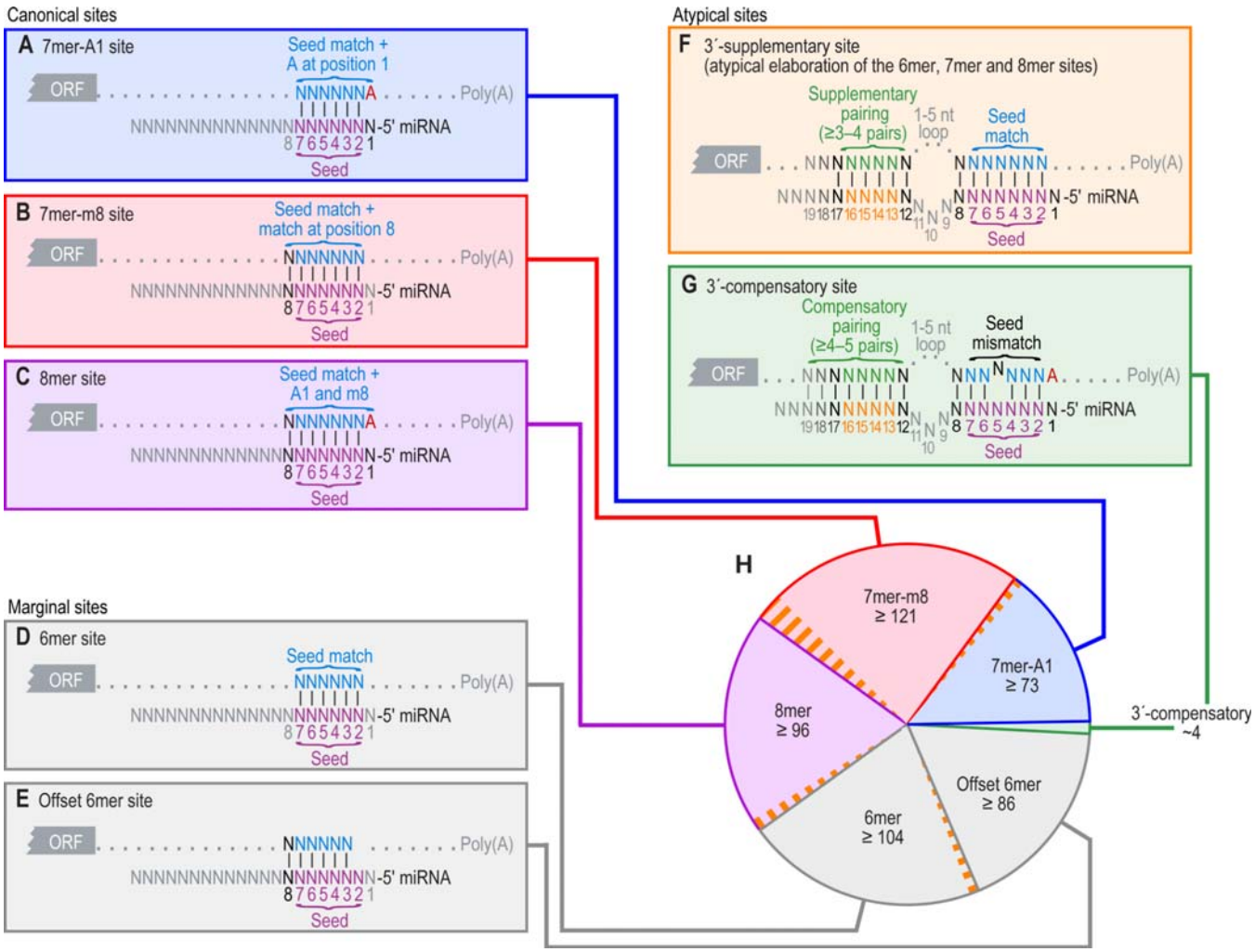

Fig. (3). miRNA target sites (Bartel. 2009)

Involvement of miRNAs in cancer:

Deregulation of miRNAs has been demonstrated in several types of human cancers (Calin, 2004). Several which are deregulated in cancer system are described in table-1.

Table 1. Deregulated miRNAs in cancer

miRNA	Targets	Aberrant expression	Roles
miR-21	PTEN	Upregulation	EMT upregulation
miR-200 Family	ZEB ½	Downregulation	EMT upregulation

miR-34a	VEGF	Induced Upregulation	Angiogenesis Downregulation
miR-10b	Syndecan-1, HOXD-10	Upregulation	Increased migration/invasion
miR-31	ITG5, RhoA, MMP16, FZD, RDX	Upregulation	Decreased migration/invasion
miR-125b	STARD13	Upregulation	Increased migration/invasion
miR-193b	Upa	Upregulation	Increased migration/invasion
miR-194	Talin-2	Upregulation	Decreased migration/invasion (in HER-2 over expressing cells)
miR-632	DNAJB6	Upregulation	Increased migration/invasion
miR-137	ERR	Upregulation	Decreased migration/invasion
miR-224	RKIP	Upregulation	Increased migration/invasion
miR-218	ROBO-1	Upregulation	Increased migration/invasion

(Adopted from Harquail et al. 2012)

Long non coding RNAs:

LncRNAs have insignificant open reading frame. It has minimum 200nt base pair length. Capping and polyadenylation also occurs like that of mRNA (Carninci et al. 2005). Both single-stranded and double-stranded long ncRNAs are reported. Double stranded RNA is formed by the transcription of Sense and antisense strands of Alu repeats (Wang et al. 2008) but its functional consequence is still remains unclear. A possible explanation for double

strandedness of ncRNAs is that the double-stranded ncRNA might not bind a target molecule and helps in repression of the ncRNA function. LncRNAs are known to be localized in nuclei and some are localized both in nuclei and cytoplasm (Imamura et al. 2004). There are no such computational methods for the detection lncRNA.

Classification of lncRNAs:

There are five broad categories of lncRNAs such as sense, antisense, bidirectional, intronic, and intergenic (Ponting etal, 2009) based upon the proximity between neighbouring transcripts.

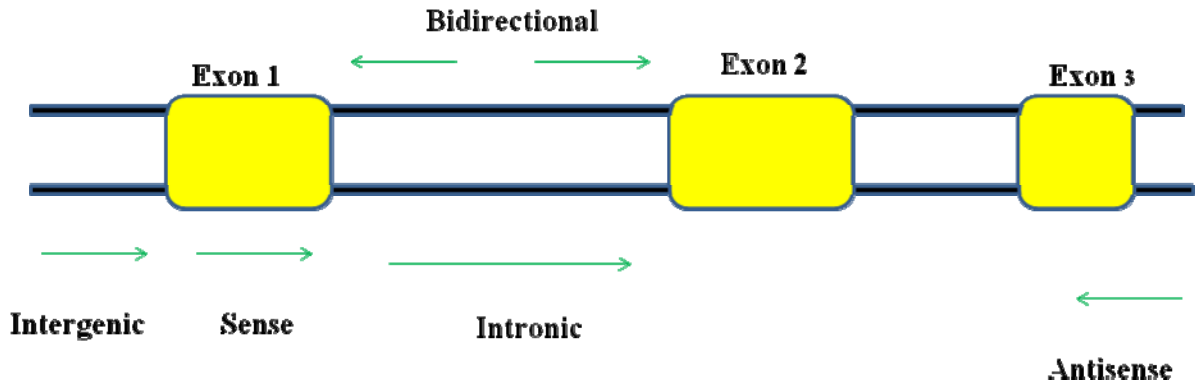

Fig. (4). Types of lncRNAs

Role of lncRNA in cancer:

A number of dysregulated lncRNAs have been detected from Genome-wide transcriptomic analyses in cancer tissues (Matouk et al. 2009). It is found that some lncRNAs promote abnormal cell proliferation and tumorigenesis, whereas some prevent the formation of tumor. The challenging thing is to define which one is cancerous and which one is not. Table-2 shows different types of deregulated lncRNA in cancer system.

Table 2. LncRNA in different cancer systems and its expression

lncRNA	Expression significance
MALAT1	Upregulation in hepatocellular carcinoma, lung cancer, breast cancer, pancreatic cancer, colon cancer, prostate cancer, and so on
HOTAIR	Upregulation in breast cancer, hepatocellular carcinoma

ANRIL	Upregulation in prostate cancer, coronary heart disease, cerebrovascular disease, diabetes, melanoma, and gliomas
HULC	Highly expressed in hepatocellular carcinoma
lincRNA-p21	Downregulation in lung cancer
GAS5	Downregulation in breast cancer
MEG3 CRNDE	Downregulation in pituitary adenomas Upregulation in Colorectal cancer and Glioma
PTENP1	Downregulation in Prostate cancer

Role of lncRNA in breast cancer:

Breast cancer is a very familiar cancer in female but still there are many obstacles for the treatment of breast cancer due to lack of understanding at the molecular level. In this regard, RNA molecule is the best choice for detection of mechanism. This RNA molecule can identify a large set of unidentified or junk RNA called long non coding RNA in human breast cancer. LncRNAs are known to be deregulated in many kinds of cancer and helps in maintaining cancer cell characteristics. It is very interesting fact that some breast cancers respond estrogens while others dependent on growth factors. Regarding this fact, there is some predictions. LncRNAs (1) are principal regulators of estrogen- or growth factor-dependent breast cancer cell (2) will indicate the type or stage of breast cancer, and (3) are very much useful therapeutic targets for treatment of breast cancer. Above statements can helps to determine specific role of lncRNA in breast cancer and also helps in probing the molecular mechanism behind it. The first human long non-coding RNA is H19 gene (Rainier et al. 1993) which is aberrantly expressed in breast cancer (Lottin et al. 2002). LncRNA *HOTAIR (*2.2kb*)* expression is strongly associated with metastases of breast cancer patients, and expression levels correlate positively with a poor outcome (Gupta et al, 2010). HOTAIR helps chromatin modification. LncRNA BCYRN1 (BC 200) is highly upregulated in breast, ovarian, cervical, lung and other cancer (Chen et al. 1997). MALAT1, a 7kb lncRNA, has role in non-small cell lung cancer, (Ji et al. 2003) and is also deregulated in breast cancers (Guffanti et al. 2009).

HULC and cancer:

HULC is an lncRNA having length 500nt. HULC refers to highly upregulated in liver cancer. It has been reported that HULC act as endogenous sponge and down regulates a series of miRNA like miR-372, miR-613 (Jiayi et al. 2010) due to its promoter activity. Repression of miR-372 leads to upregulation of its corresponding gene called PRKACB. Later on HULC is also over expressed in gastric cancer (Zhao et al. 2014). But till now it is not reported in breast cancer. So it will be a newer approach if it is still remain upregulated in breast cancer.

miR-30a-3p and cancer:

miR-30a-3p is downregulated in Hepatocellular carcinoma and it acts as a tumor suppressor in vitro (Wang et al. 2013). It is also down regulated in colorectal cancer (Ma et al. 2012). It is not reported in breast cancer yet.

OBJECTIVES:

Objective 1:

HULC, as the name suggests, found in liver cancer. We want to know whether it is significantly expressed in other systems like breast cancer,

Objective 2:

We want to explore why this lncRNA is upregulated and what is its impact on oncogenesis of breast cancer.

Objective 3:

We are curious to know whether conventional targeting or ceRNA is involved in miRNA-lncRNA interaction is operational in breast cancer contributing to its oncogenesis.

Objective 4:

Microarray expression analysis for identification of differentially expressed miRNA in breast cancer.

Objective 5:

RNA seq analysis for identification of differentially expressed lncRNA in breast cancer.

Objective 6:

Identification of novel lncRNA–miRNA target pairs involved in breast cancer & experimental validation of both miRNA and lncRNA expression by RT-PCR in MDA-MB-231 breast cancer cell lines and HaCaT control cell line.

MATERIALS AND METHODS:

1.1 Data retrieval from microarray data.

For genome wide analysis of miRNA and mRNA of normal and diseased sample gene expression data was taken. From Gene Expression Omnibus (GEO) database, the gene expression data were retrieved. GEO a public data repository and data retrieval system where next-generation sequencing, microarray and other forms of high-throughput functional genomic data are freely available, submitted by scientific community. This database supports MIAME-compliant data submissions. A collection of web-based interfaces and applications are available to provide help users query and download experiments and curated gene expression profiles stored in GEO. The GEO data contains raw microarray data which involves images that are to be converted into gene expression matrices; rows represent genes, columns represent various samples such as tissues or experimental conditions, and numbers in each cell illustrate the expression level of the particular gene in the particular sample. Analysis of the matrices can be further done in order to extract any biological process and its understanding. Platform describes the list of features on the array (e.g., cDNAs, oligonucleotides, etc.). There is an importance of using different platforms as because of the diversity of technical and analytical sources that can affect the results of an experiment. Therefore, a single platform may be inefficient for comparison among experiments. So, large-scale comparison studies involving microarrays can be done using various platforms for optimum reproducibility measurements.

There are 2 ways by which the GEO data may be retrieved

- GEO data may be retrieved by querying GDS (GEO datasets), gene profiles and GEO accession number.
- GEO data can be accessed directly on the web by browsing through datasets or GEO accessions options available with respect to individual platform, sample and series. Related records are intra-linked on the GEO sites such that one may conveniently navigate to the associated platform, series, and sample and GEO dataset records.

Fig. (5) GEO datasets

1.2 Analysis through GEO data base:

- Go to main page of GEO data base.
- Type mRNA expression in breast cancer AND Homo sapiens in data sets option.
- Then click on GO.
- Then results appear which contain the list of microarray experiments. From that each GSE is checked for expression profile analysis having platform Agilent-031181 Unrestricted_Human_miRNA_V16.0_Microarray 030840 (miRBase release 14.0 miRNA ID version),the expression is without induce of any drug.
- From that **GSE38867 is** taken for analysis.
- From that 5 controls (normal) and 3 tests (disease) samples were taken. To reduce the error triplicate should taken.

Table 3. Shows normal and test samples.

CONTROL	TEST
GSM951056	GSM951062
GSM951068	GSM951066
GSM951048	GSM951054
GSM951064	
GSM951060	

- From that, a raw file was down loaded, it was unzipped and for further analysis it was uploaded in GeneSpring software.

1.3 Analysis of gene expression data:

- Expression data is analyzed by using GeneSpring software.

GeneSpring software

GeneSpring is software which provides statistical tools for visualization and data analysis. For understanding of Transcriptomic, Metabolomics, Proteomics and NGS data within a biological context, GeneSpring is very useful. It also helps in expression analysis. This software is a key component of systems biology research, involves producing heterogeneous data that measure various biological entities and events such as variation in expression of microRNA and mRNA exon splicing, DNA structure, proteins and metabolites, which helps to understand underlying mechanism of diseases. It also allows researchers to analyse, compare different signalling pathway.

Other applications of GeneSpring include:

- Gene expression analysis of microarray platforms like Affymetrix, Agilent and Illumina.
- Analysis of Real-time PCR data.
- GEO datasets.

Data analysis using GeneSpring

For miRNA expression analysis:

Procedure

- The raw data files from GEO database were downloaded as a zip file.
- The files then were unzipped, extracted and renamed according to the convenience.

- Firstly, a new project was selected where many platforms came. From that, a desired platform type was selected.
- Then the name of the project was given.
- A new experiment was set by setting experiment name and experiment type.
- Then guided work flow was chosen as work flow type.
- Then data is loaded and the desired technology is selected i.e. Agilent-031181 Unrestricted_Human_miRNA_V16.0_Microarray 030840 (miRBase release 14.0 miRNA ID version).Then files were chosen as 3 replicates i.e. Control and Test files.
- Then by normalizing the noise signal the intensity map could be seen.
- Then **Experimental setup** was done by following methods.
    - First **experimental grouping** was done by adding parameters to **Average.**
    - Then **Average over replicates in condition** was chosen in case of **Create Interpretation.**
- Then **quality control** is done by following methods.
    - First a **Filter Probe sets by errors** was set.
    - Then the **coefficient of variation** was chosen **< 50%.**
    - Then **Output views of filter** were chosen.
    - Then the **entity list** was saved.
- Then **Analysis** was done by following method
    - First **Statistical Analysis** was chosen.
    - After that **Entity list** was Filtered on **error <50.**
    - Then the **Interpretation** was chosen **Average**.
    - **Condition 1** and **Condition 2** for **Select test**.
    - **T-test is unpaired chosen** for 2 sets of data. If more than 2 sets be present then **Anova** should be chosen.
    - Then the **p- value computation** was chosen **Asymptomatic.**
    - Then the **Number of permutation** should be **100**.
    - Then **multiple testing corrections** was chosen **Benjamin Hochberg FDR.**
    - Then in **Results** cut-off should be chosen **≤ 0.05** which was present by default.
    - Then the **entity list** was saved.
- Then the **Fold Change** was chosen and done by following these **Input parameters** and **pairing options.**
    - The **Fold change results** cut-off should be **>=2.0.**

- Then **Clustering** was done by following methods:
  - First **Input parameters** were taken.
  - Then in **Entity list**, Fold change **>=2.0** was chosen**.**
- Then the **Interpretation** was chosen **all samples**.
- The **Cluster algorithm** was chosen **Hierarchical.**
- The **Cluster** was done on **Conditions**
- Then the **Distance metric** was chosen **Pearson uncentered**.
- The **Linkage rule** was chosen Average**.**
- Then Result **file** was saved.
- After that **Result interpretation** was done by following method
- Then GO analysis by choosing output views.
- **Export** of **Fold change data** was done by adopting the following method:
  - **Entity list** was exported.
  - Normalization of the **signal values** was done.
  - Then - **miRNA** was selected in **selected items**.
  - Then **Interpretation** of **All samples** was done.
- Then it was saved as- Tab separated file (.txt)
- Then we have to see how the control varies from test sample.
- To see how the control varies from test sample clustering was done by using **Cluster3** and **java treeview software** was used.

RNA-seq analysis of lncRNAs:

The transcriptional profiling data of breast cancer was obtained from GEO database having the accession number GSE28866.The retrieved dataset comprised of transcriptional profiles for six normal breast tissue samples and six breast cancer tissue samples. We have used the RNAseq protocol as follow.

System Specification

The work was carried out on ProLite HP server with the below specification.

1. Operating system: CentOS Release 6.5 (Final).
2. Kernel-2.6.32-431.el6.x86_64

3. GNOME 2.28.2
4. Hardware: Memory: 15.4 GiB
5. Processor: Intel(R) xeon(R) CPU E5-2420 0 @ 1.90GHz

Software Requirement:

1. SRA tool kit
2. FastQC
3. FASTX
4. SAMtools
5. BOWTIE
6. GeneSpring NGS 12.6

Quality check and Filtration:

SRA tool kit (fastq-dump) was used to convert SRA files into fastq files. The qualities of the raw reads were found to be reliable predicted by FastQC and the illumina adapters present in the reads were filtered out using FASTX.

Command to run fastqc**:**

fastqc -t SRR191639.fastq

Command to run fastx_clipper**:**

fastx_clipper –a illumina_adapter.fasta -i SRR191639.fastq –o SRR191639_clipped.fastq

Reference Genome Mapping:

The filtered reads were mapped to the human reference genome (hg18) using BOWTIE and the aligned files were saved in sam format.

Command to run BOWTIE**:**

bowtie -q -m 1 –v 3 --sam --best --strata hg18 SRR191639_clipped.fastq - > SRR191639_mapped.sam

File Format Conversion and sorting:

SAMtools was used to convert the file into BAM format followed by sorting the reads in aligned files and index generation.

Command to run SAMtools:

samtools view -b -S -o SRR191639_mapped.bam SRR191639_mapped.sam

samtools sort SRR191639_mapped.bam SRR191639_sorted.bam

samtools index SRR191639_sorted.bam

Quantification and Differential Expression Analysis:

Transcript assembling was carried out using GeneSpring NGS 12.6 refereeing RefSeq as the annotated gene map. Differential expression analysis was performed using DESeq (Bioconductor Package) inbuilt in GeneSpring NGS 12.6.

Retrieval of miR-30a-3p from RNAhybrid:

All total 218 differentially expressed miRNA was taken from microarray dataset. Sequences of all the miRNA was retrieved from miRBase software. This data is then hybridised with HULC by RNAhybrid software. By doing this we got miR-30a-3p.

RNAhybrid software:

RNAhybrid is a tool for finding the minimum free energy hybridisation of a long (target) and a short RNA (query). The hybridisation is performed in a kind of domain mode, i.e. the short sequence is hybridised to the best fitting part of the long one. The tool is primarily meant as a means for microRNA target prediction. The web form of the online version of RNAhybrid offers a various number of options (shown below) to run RNAhyrbid successfully.

Target RNA:

Each of the target sequences in target file is subjected to hybridisation with each of the queries (which either are from the query file or is the one query given on command line; see -q below). The sequences in the target file have to be in FASTA format, ie. one line starting with a > and directly followed by a name, then one or more following lines with the sequence itself. Each individual sequence line must not have more than 1000 characters. If no -t is given, either the last argument (if a -q is given) or the second last argument (if no -q is given) to RNAhybrid is taken as a target.

Query RNA:

see target RNA description

Hit per target (-b)

Maximal numbers of hits to show. Hit number hits with increasing minimum free energy (reminder: larger energies are worse) are shown, unless the -e option is usedand the energy cut-off has been exceeded (see -e option below) or there are no more hits. Hits may only overlap at dangling bases (5' or 3' unpaired end of target).

Compact output (-c) (optional):

Produce compact output. For each target/query pair one line of output is generated. Each line is a colon (:) separated list of the following fields: target name, query name, minimum free energy, position in target, alignment line 1, line 2, line 3, line 4.

Helix constraint (-f) (optional):

Forces all structures to have a helix from position from to position to with respect to the query. The first base has position 1.

We have given 2-7 bp as helix constraint.

Maximal internal loop size (-u) (optional):

The maximally allowed number of unpaired nucleotides in either side of an internal loop.

Maximal bulge loop size (-v) (optional):

The maximally allowed number of unpaired nucleotides in either side of an internal loop.

Energy cut off (-e) (optional):

Hits with increasing minimum free energy (reminder: larger energies are worsening) less than or equal to energy cut off are shown, unless the –b option is used and the number already reported hits has reached the maximal hit number. Hit may only overlap at dangling bases (5' or 3' unpaired end of target).

We have given -15 as energy cut off.

Experimental Validations:

Animal Cell culture:

MDA MB 231, a human breast carcinoma cell line, was taken from National Centre for Cell Science (NCCS), Pune, India. Before culturing these cell lines are maintained in 20° c in carbon dioxide incubator. Media used for these cell cultures is **DMEM.**

MDA-MB-231:

It is breast cancer cell line. Requires DMEM medium for growth.

HaCaT:

It is the control cell line which also grows in DMEM media.

Steps for cell culture:

Cell thawing:

1. Water was heated up to 37'c and within which the cells were plated. A vial of frozen cells were removed from liquid nitrogen and placed in the water bath until it was thawed.
2. Outside of the vial was washed with 70% alcohol.
3. Then the thawed cell specimen was pipetted into cell culture vessel containing growth medium. The vessel was swirled gently to mix the cell with the medium.
4. Sometimes removal of DMSO is necessary for suspension cell, primary cells and sensitive cell types. For such cell types, the thawed cell suspension was pipetted into a sterile centrifuge tube containing medium. Centrifugation was done at 200g for 2 minute. Then supernatant was aspirated and cells were resuspended into fresh medium.
5. Cells were incubated overnight under their usual growth condition.
6. In the next day, growth medium was replaced.

   Trypsinizing cells:

   Trypsinization is a technique which uses the proteolytic enzyme trypsin to detach adherent cells from the surface of a cell culture vessel. This procedure is followed whenever the cells need to be harvested.
7. The medium was aspirated and then discarded.
8. The cells were washed with PBS (phosphate buffer saline).
9. 1x trypsin EDTA was added to the solution.

10. Then the flask was placed in a CO2 incubator at 37'c for 1-2 min.
11. Flask was removed from incubator and then it was tapped with palm to assist detachment.
12. Once discarded the cells were then resuspended in growth medium containing serum. The serum inactivates trypsin activity.
13. The cells were gently pipetted up and down for proper mixing.
14. Then proceed as required. (e.g-with freezing, nucleic acid isolation.)

RNA Isolation:

The kit used for miRNA is called mir*V*ana from QIAGEN.

1. The cells were trypsinized and collected. Then it was washed with PBS and pellet was collected, and then was placed in ice.
2. PBS was removed. Lysis binding solution was added if cells were 100s in number-300 µl and if cells were 1000s in number-600 µl. Then vortexing was done vigorously to make homogenous lysate.
3. 60 µl of miRNA homogenate additive was added to the cells. Then it was mixed well by vortexing for several times and then it was remained on ice for 10 minutes.
4. 1 volume of acid-chloroform was added and vortexing for 30-60 sec. Centrifugation was done at 10,000gnto separate aqueous and organic phase. After centrifugation, the interphase should be compact if it is not repeat centrifugation. The aqueous phase was removed without disturbing the lower phase and was transferred it to a fresh tube.
5. 1.25 volume of room temperature 100% ethanol was added to the aqueous phase.
6. Filter cartridge was placed into one of the collection tube. Lysate ethanol was pipetted and mixed into the filter catridge (up to 700 µl, add step wise thereafter). Centrifugation was done for 15 sec at 10,000 rpm. Flow through was discarded.
7. 700 µl of miRNA wash solution 1 was added to filter catridge. Centrifugation was done for 5-10 sec. Flow through was discarded.
8. 500 µl of buffer RPE and centrifugation was done for 15sec at ≥8000×g.. Flow through was discarded
9. 500 µl of buffer RPE and centrifugation was done for 2 min at >8000xg. Flow through was discarded at ≥8000×g.

10. The RNeasy spin column was placed in the new 1.5 ml collection tube. 30-50 µl RNasefree water was added directly to the spin column membrane. The lid was closed and centrifuged for 1min at ≥8000×g to elute the RNA.
11. If the expected RNA yield is >30 µg, then step 7 was repeated using another 30-50µl of RNase- free water, or using the eluate from step-7. The collection tubes were reused from step-7.
12. The purity and yield of RNA yield was measured by **Eppendorf NanoDrop**. It is a cuvette free spectrophotometer which eliminates the need for other sample containment devices and allows for clean up in seconds. It measures 1 µl samples with high accuracy and reproducibility. The full spectrum (220nm-750nm) spectrophotometer utilizes a patented sample retention technology that employs surface tension alone to hold the sample in place. A 1 µl sample is pipetted onto the end of a fiber optic cable (the receiving fiber). A second fiber optic cable (the source fiber) is then brought into contact with the liquid sample causing the liquid to bridge the gap between the fiber optic ends. The gap is controlled to both 1mm and 0.2 mm paths. A pulsed xenon flash lamp provides the light source and a spectrometer utilizing a linear CCD array is used to analyze the light after passing through the sample. The instrument is controlled by PC based software, and the data is logged in an archive file on the PC.

cDNA synthesis:

cDNA synthesis was carried out using SuperScript First-Strand Synthesis System for RT-PCR by Invitrogen using oligo dT primers.

Procedure:

1. Each of the components were mixed and briefly centrifuged before use.
2. For each reaction, the following in a sterile 0.2 or 0.5ml tube was combined

Table 4. Components using in cDNA synthesis.

Components	**Amount**
RNA	4 µl
10 mM dNTP mix	1 µl
Primer (0.5µg/µl oligo (dT)12-18 or 2µM gene specific primer)	1 µl

DEPC treated water	2 µl

4. The RNA/primer mixture at 65·c for 5 minutes was incubated, and then placed on ice for at least 1 minute.

5. In a separation tube, the following 2X reaction was prepared by adding each component in the indicated order.

Table 5. Different components used in cDNA synthesis.

Components	**1RXn**	**10 RXns**
10X RT buffer	2 µl	20 µl
25mM Mgcl2	4 µl	40 µl
0.1M DTT	2 µl	20 µl
RNase out TM (400/ µl)	1 µl	10 Ml

7. 9µl of the 2X reaction mixture was added to each RNA/primer mixture from step3, mixed gently and collected by briefly centrifuge.
8. It was incubate at 42°c for 2 minutes.
9. 1µl of super script TM II RT was added to each tube.
10. It was incubate at 42°c for 50 minutes.
11. The reaction was terminated at 70·c for 15 minutes. Chilled on ice.
12. The reaction was collected by brief centrifugation. 1µl of RNase H was added to each tube and incubated for 20minutes at 37·c. The reaction was used for PCR immediately.

Quantitative Real Time RT-PCR Analysis:

Real-time PCR is the continuous collection of fluorescent signal from one or more polymerase chain reactions over a range of cycles. Quantitative RT-PCR involves the conversion of the fluorescent signals from each reaction into a numerical value for each

sample. Fluorescent marker is used which binds to the DNA. Therefore, as the number of gene copies increases during the reaction so the fluorescence intensity increases. This is advantageous because the rate of the reaction and efficiency can be seen. Intercalating fluorescent dyes (e.g. SYBR green) are the simplest and cheapest way to monitor a PCR in real-time. The SYBR green dye fluoresces only when bound to double-stranded DNA. The major disadvantage of using a dye such as this is the lack of specificity.

MiRNA primer sequence was designed by us using tools like Amplifx (1.5.4) and ApEA plasmid editor. This sequence was ordered from SIGMA GENOSYS. All the primers were desalted and UV absorbance was used to assess the quality of primer synthesis.

To perform PCR using RNA as a starting template which must first be reverse transcribed into cDNA in a reverse transcription (RT) reaction, where the cDNA is used as template for real-time PCR with gene specific primers.

Table 6. Primer name and sequence with length

Primer name	5'<----sequence--->3'	Length
beta actin F	CATGTACGTTGCTATCCAGGC	21
beta actin R	CTCCTTAATGTCACGCACGAT	21
HULC F	ATCTGCAAGCCAGGAAGAGTC	21
HULC R	CTTGCTTGATGCTTTGGTCTGT	22
miR-30a-3p F	GCCCTTTCAGTCGGATGTTTGCAG	24

Quantification of HULC:

Real-time PCR was carried out in Eppendorf Masterplex Real Time PCR.

1. Total master mix was prepared for 3 genes; one house keeping gene, HULC, and two samples: HaCaT and MDA-MD-231.
2. Following components were required for quantification.
3. cDNA was diluted into 1:20 ratio concentration
5. The following mixture was taken as follows for preparation

Table 7. Components for lncRNA, HULC:

Component	Amount
SYBR Green Mix	65 µl
cDNA stock (cDNA: dH2O [1:20])	39 µl

primer pair mix (3.25 µl each primer)	13 µl

6. Master mix contains all the components except cDNA. 13 µl of water was added to the master mix and the mix was divided into two aliquots.

7. To each aliquot, forward and reverse primer (3.25 µl each) was added. This mix was again divided into two aliquots.

8. Then to each reaction, 9. 75 µl of cDNA was added.

10. Each reaction mix was divided into replicates.

11. Then it was in RT-PCR. The experiment was set up and the following PCR program was made on. A copy of the setup file was saved and all PCR cycles were deleted. The threshold frequency taken was 33%. The cycle temperatures taken were as follows:

Table 8. Cycle temperature and time for qRT-PCR

STAGE	**TEMPERATURE (°C)**	**TIME**	**CYCLE**
Stage 1	95	20sec	1
Stage 2	95 55 68	15sec 15sec 20sec	40
Stage 3	95 60 95	15sec 15sec 15sec	1

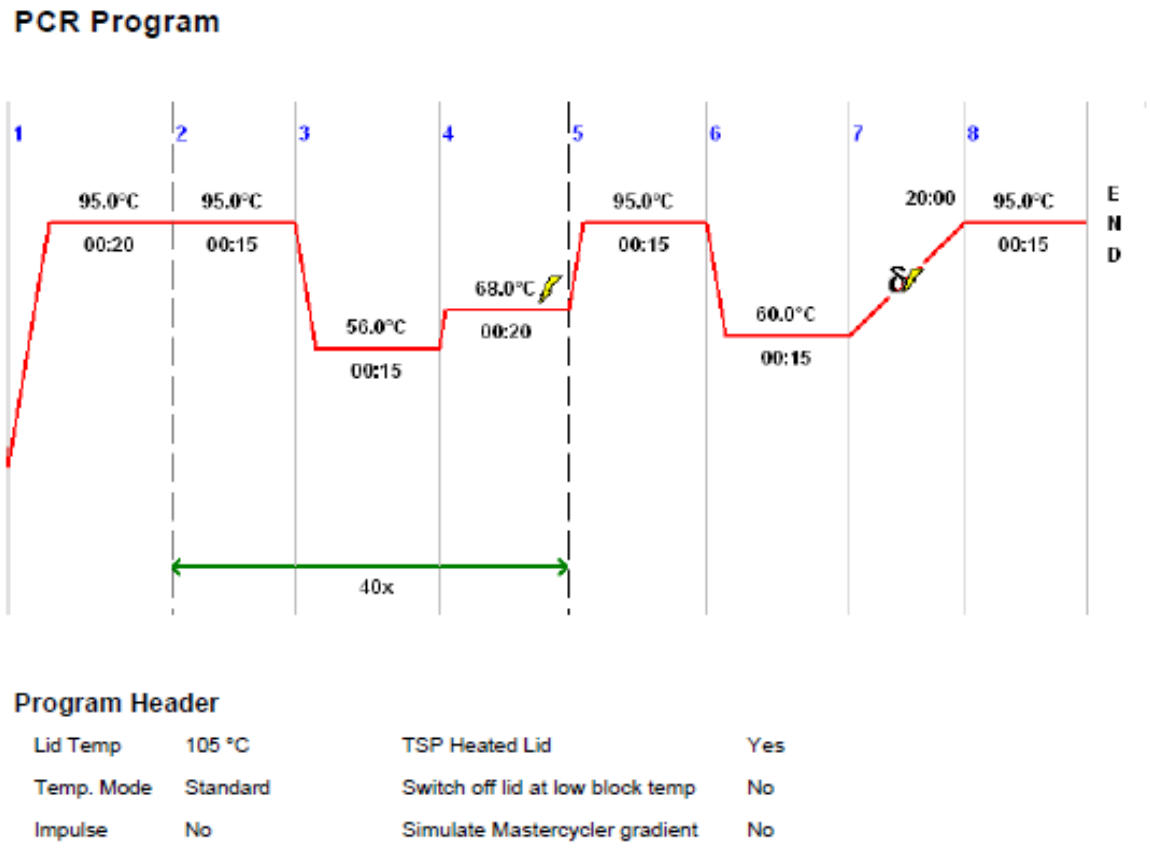

Fig. (6). Cycle temperature and time for qRT-PCR

Quantification of miR-30a-3p:

All the above procedure was also same for miRNA except the following components.

Table 9. Components for miRNA miR-30a-3p

Component	**Amount**
SYBR Green Mix	65 µl
cDNA stock (cDNA: dH2O [1:20])	13 µl
primer pair mix (3.25 µl each primer)	5.2 µl
ROX	0.52 µl

RESULTS AND DISCUSSIONS:

miRNA expression result from microarray analysis:

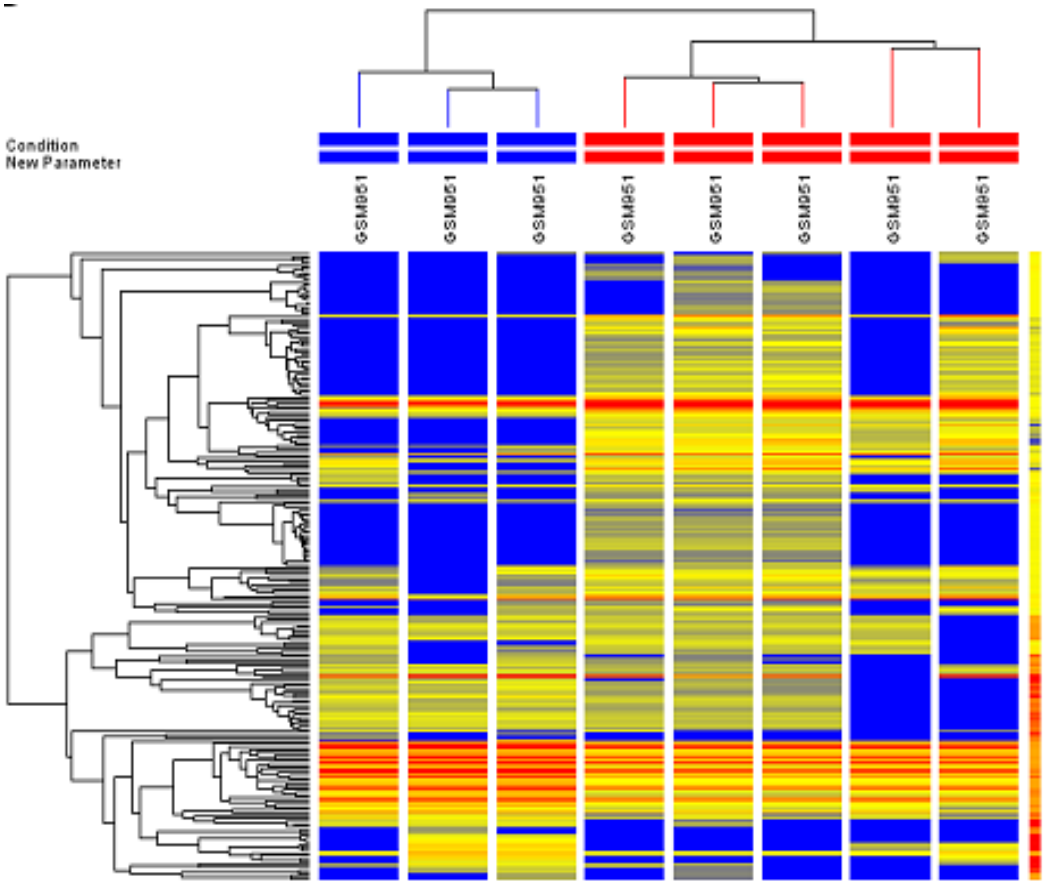

Fig. (7). Clustering of miRNA

From this figure, different expression pattern of miRNAs were seen. Blue colour shows down-regulation, yellow shows normal and red colour shows up-regulation. Based upon their similarity in expression, miRNAs clustered together.

RNA-seq result:

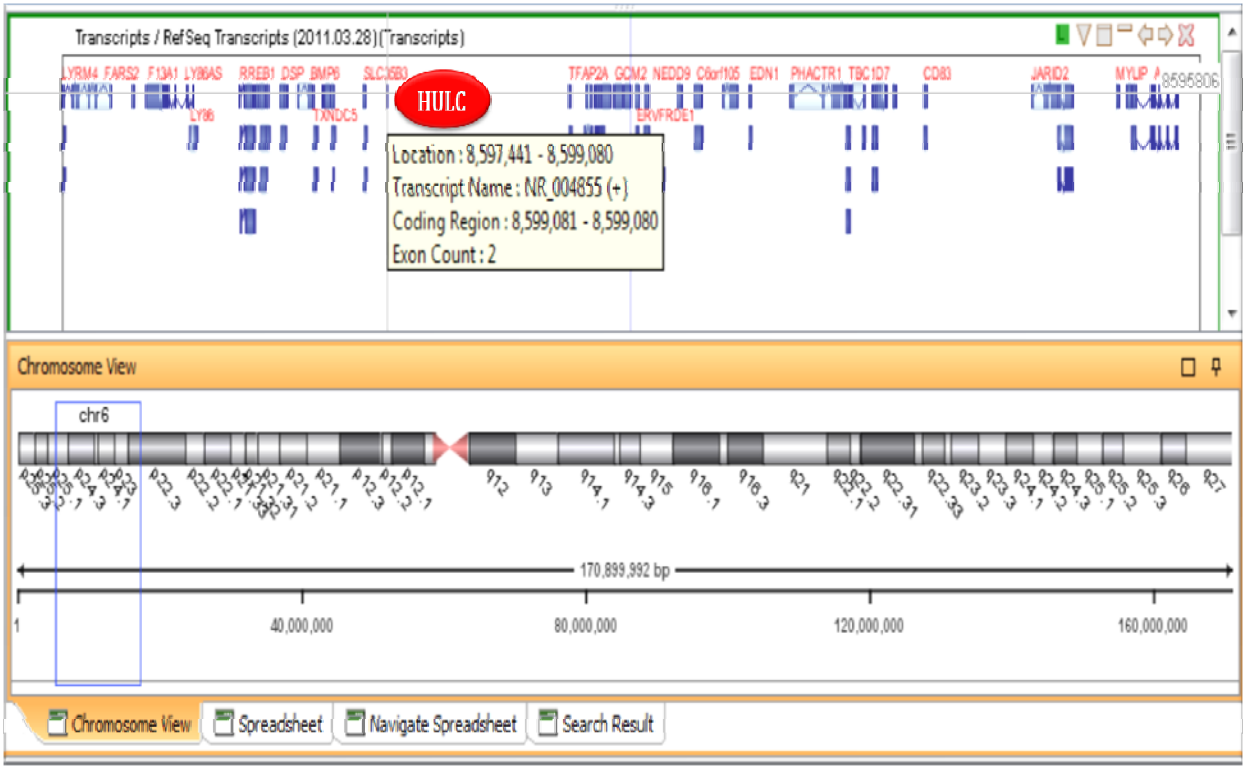

Fig. (8). Location of HULC in genomic DNA

Above figure shows location of HULC in chromosome 6. We choose HULC for our experiment because it is highly expressed and yet not validated in breast cancer.

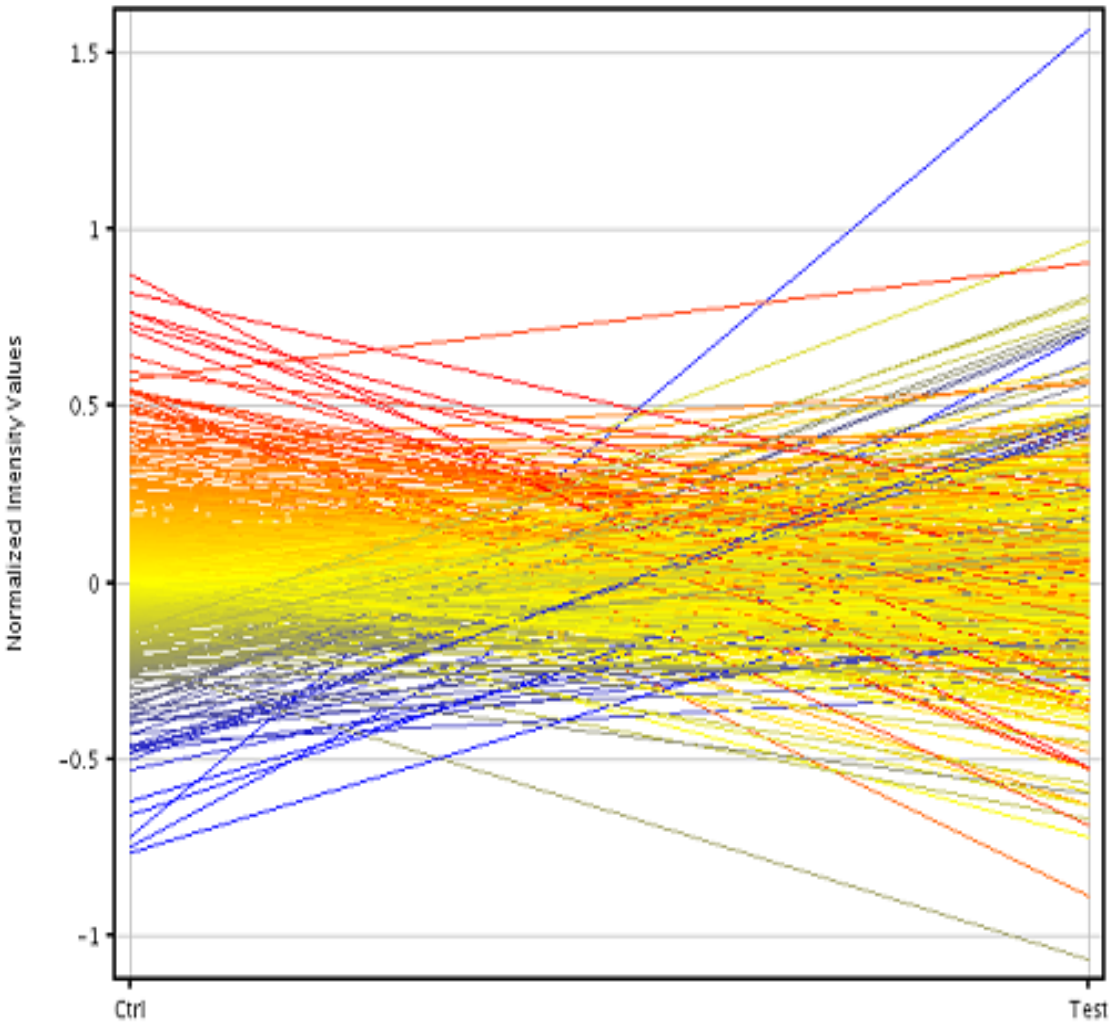

Fig. (9). Normalized intensity values of genes

Before quality check each entity should be normalized to reduce the systematic variation. Above figure shows the expression value of different genes.

RNAhybrid result:

We took 218 miRNA for interaction with HULC. Among this, 9 miRNA was complementarily bind with HULC. From these 9 miRNA, miR-30a-3p was selected as it has complete Watson-Crick base pairing in seed region as well as in other region and energy cut off is also more.

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-10b-3p
length: 22

mfe: -16.0 kcal/mol
p-value: undefined

position  90
target 5'   G                  C 3'
             CCCC         AAUCUG
             GGGG         UUAGAC
miRNA  3' UAA    AUCUUAGC      A 5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-1273e
length: 22

mfe: -17.9 kcal/mol
p-value: undefined

position  262
target 5' A   AAG A             G 3'
           CCA   C       UCAAGCAA
           GGU   G       AGUUCGUU
miRNA  3' A   GAA GACCCA          5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-141-3p
length: 22

mfe: -15.0 kcal/mol
p-value: undefined

position  310
target 5' U       UAA   ACACAAAUUA      C 3'
           UAUUUUU   GGA           AGUGUU
           GUAGAAA   UCU           UCACAA
miRNA  3' G       UGG   G               U 5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-149-5p
length: 23

mfe: -27.4 kcal/mol
p-value: undefined

position  73
target 5'    C          AGCAA  CCCCAAUCUGCA       A 3'
              GUGAGGAUAC     GG             AGCCAGG
              CACUUCUGUG     CC             UCGGUCU
miRNA  3' CCCU                                    5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-27a-3p
length: 21

mfe: -20.4 kcal/mol
p-value: undefined

position  135
target 5' U  A      CUGAUCG   A AUUUCAACCUCCAGA       U 3'
           GC GGAACU       UGG C                ACUGUGA
           CG CCUUGA       AUC G                UGACACU
miRNA  3'                      G                       U 5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-30a-3p
length: 22

mfe: -21.2 kcal/mol
p-value: undefined

position  403
target 5' A  AUG     AUAU    AAA       C 3'
           GC   GCAAA    CAUC   ACUGAAA
           CG   CGUUU    GUAG   UGACUUU
miRNA  3'    A              GC        C 5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-4312
length: 19

mfe: -27.2 kcal/mol
p-value: undefined

position  73
target 5'    C UG    UACA        C 3'
              G  AGGA    GCAAGGCC
              C  UCCU    UGUUCCGG
miRNA  3' ACCC UG                 5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-563
length: 19

mfe: -18.4 kcal/mol
p-value: undefined

position  138
target 5' A    CUCUGAU    GACAUU       C 3'
           GGAA       CGUG      UCAACCU
           CCUU       GCAU      AGUUGGA
miRNA  3' C    U          AC             5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-563
length: 19

mfe: -18.3 kcal/mol
p-value: undefined

position  320
target 5' A    CACAAAUUAA           G 3'
           GGAA          GUGU UCAACCU
           CCUU          CAUA AGUUGGA
miRNA  3' C    UG            C         5'
```

```
dataset: 1
TARGET : NR_004855.2|HULC
length: 500
MIRNA  : hsa-miR-664a-5p
length: 24

mfe: -22.8 kcal/mol
p-value: undefined

position  92
target 5'  C      UG    A        G 3'
            CCAAUC     C   AGCCAG
            GGUUAG     G   UCGGUC
miRNA  3' UA      UAAAA GGA      A 5'
```

Experimental Validations:

RNA isolation

• **260/280 Ratio:** This ratio indicates the absorbance of DNA and RNA at 260 nm and 280 nm, which is used to assess the purity of DNA and RNA. A ratio is expected approximately 1.8 and generally accepted as "pure" for DNA; a ratio of approximately 2.0 is generally accepted as "pure" for RNA. In either case, if the ratio is significantly lower it may indicate the presence of phenol, protein or other contaminants that absorb strongly at or near 280 nm.

• **260/230 Ratio**: This ratio is used as a secondary measure of nucleic acid purity and phenolic contamination. The 260/230 values for "pure" nucleic acid are often higher than the respective 260/280 values. Expected 260/230 values are normally in the range of 2.0-2.2. If the ratio is significantly lower than expected, the presence of contaminants which absorb at 230 nm. is indicated.

• Here, we got following results of two samples; Sample1 has perfect ratio absorbance in 260nm and 280 nm wavelengths where Sample2 has less value.

• **Sample1** = 870.5 µg/ml

At (260/280) ratio = 2.07

At (260/230) ratio = 2.01

• **Sample 2** = 828.4 µg/ml

At (260/280) ratio = 2.07

At (260/230) ratio = 1.69

qRT-PCR:

qRT- PCR melting curve analysis was used to quantify nucleic acid, mutation detection and for genotype analysis.

• The melting temperature curve for the HULC and miR-30a-3p with respect to control , β-actin was observed.

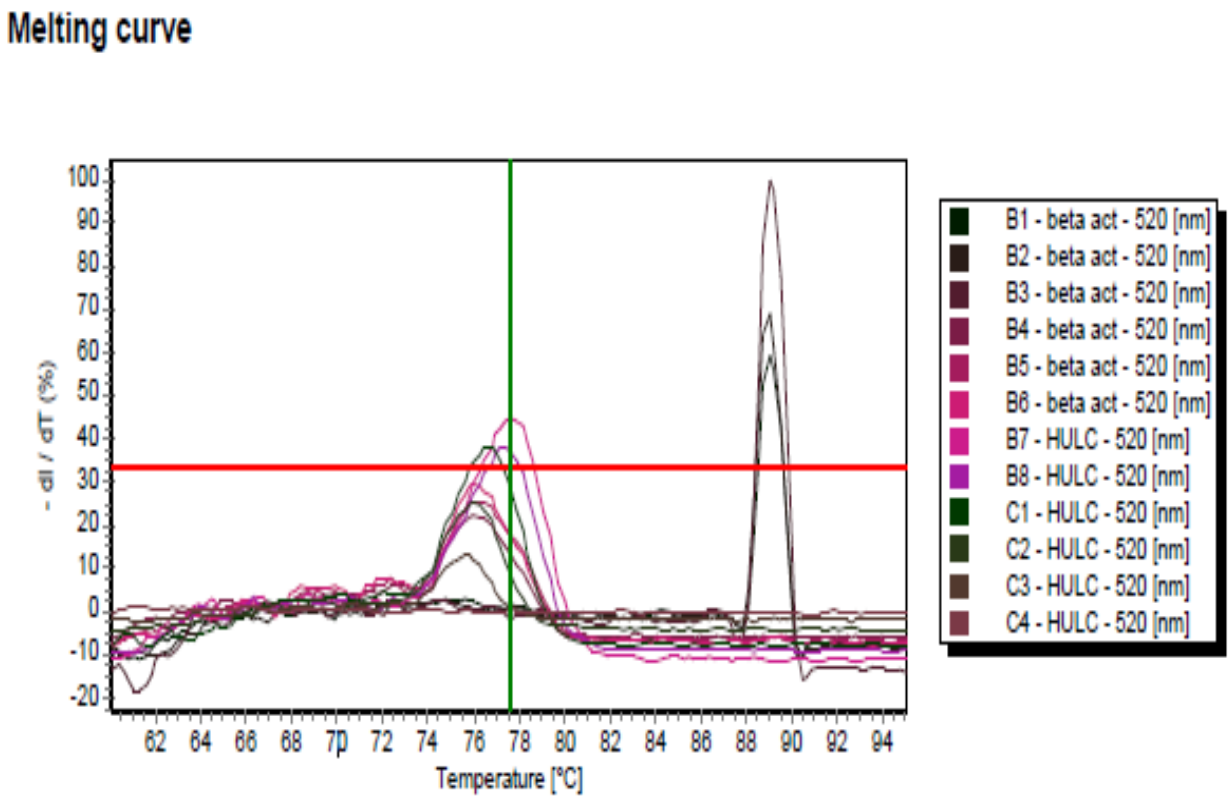

Fig. (10). Melting curve of HULC

From the qRT PCR analysis of HULC, a fine melting temperature curve was observed in comparison to the control gene, β-actin. The melting temperature of HULC is 85°c. The samples of gene were taken in triplicates. Three peaks of HULC positioned at one place were observed. The relative quantification of HULC with respect to β-actin shows high expression of HULC.

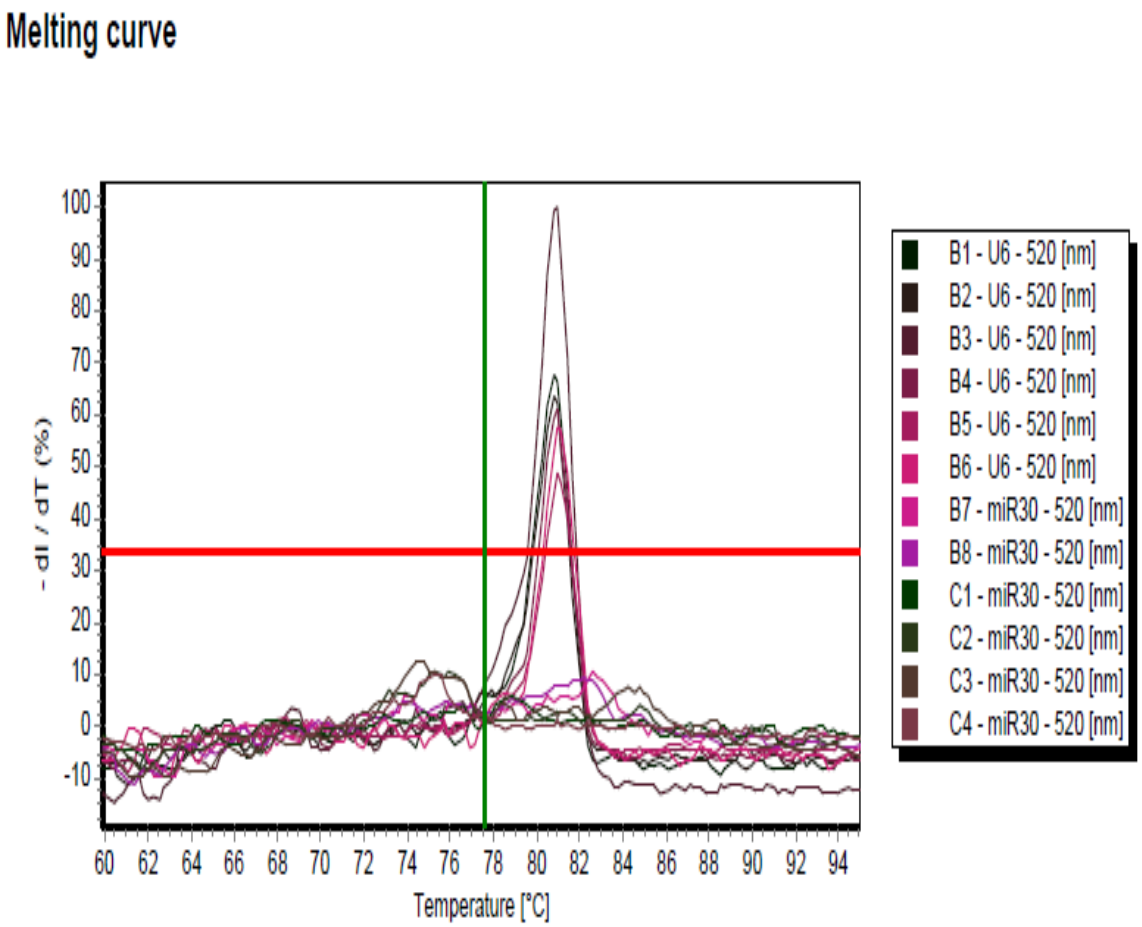

Fig. (11). Melting curve of miR-30a-3p

From the qRT PCR analysis of miR-30a-3p, a fine melting temperature curve was observed in comparison to the control gene, U6. The melting temperature of miR-30a-3p is 81°c. The samples of gene were taken in triplicates. Three peaks of miR-30a-3p positioned at one place were observed. The relative quantification miR-30a-3p of with respect to U6 shows high expression of miR-30a-3p.

CONCLUSIONS:

Our study aimed at identification of novel lncRNA-miRNA target pairs that are hypothesized to play a role in breast cancer through experimental validation of selected lncRNA and miRNA. From miRNA microarray expression analysis, we found that the total number of differentially expressed miRNA in breast cancer is 197, out of which 122 are down regulated and 75 are up-regulated. We did RNA-seq analysis where we found a huge number of lncRNAs that are expressed differentially in breast cancer. From this set of differentially expressed lncRNAs, we selected one lncRNA, HULC. HULC was reported for the first time in liver cancer and was named as "highly upregulated in liver cancer" or HULC. HULC is over-expressed in liver cancer and gastric cancer; however it was not reported in breast cancer. After selecting HULC, we hybridized this lncRNA with a down-regulated miRNA obtained from microarray data using RNAhybrid software. It was seen that HULC has complementarily with 9 miRNAs. Further screening of these HULC-miRNA pairs based on criteria like stability, more complementarily binding lead to identification of a significant miRNA-HULC pair that might have crucial role in breast cancer malignancy. None of the members of this pair are validated in breast cancer till now. Therefore, we did experimental validation of both miRNA and lncRNA by RT-qPCR in MDA-MB-231 cell line. We found high up-regulation of HULC and down-regulation of miR-30a-3p. Upregulation of HULC might be due to down-regulation of hsa-miR-30a-3p which in turn might help in histone modification which gives its oncogenic properties. We want to confirm whether one ncRNA could regulate the other.

FUTURE PROSPECTIVES:

It is a very challenging and interesting topic that how a ncRNA regulates another ncRNA. For this type of regulation, conventional way exists miRNA→RNA. But it can be assumed in reverse way i.e. RNA→miRNA through competitive endogenous RNA (ceRNA) hypothesis. If it is successful, then many hidden layers of regulation by lncRNAs will come into limelight and can help to understand the systems biology of malignancy.

REFERENCES:

Babu GR, Lakshmi SB, Thiyagarajan JA. Epidemiological correlates of breast cancer in South India. Asian Pac J Cancer Prev. 2013;14(9):5077-83.

Redig AJ, McAllister SS. Breast cancer as a systemic disease: a view of metastasis. J Intern Med. 2013 Aug;274(2):113-26.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010 Dec 15;127(12):2893-917.

Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. Mol Cancer. 2011 Apr 13; 10:38.

Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009 Jan 23;136(2):215-33.

Martin L, Chang HY. Uncovering the role of genomic "dark matter" in human disease. J Clin Invest. 2012 May 1;122(5):1589-95.

Fenoglio C, Ridolfi E, Galimberti D, Scarpini E. An emerging role for long non-coding RNA dysregulation in neurological disorders. Int J Mol Sci. 2013 Oct 14;14(10):20427-42.

Khaitan D, Dinger ME, Mazar J, Crawford J, Smith MA, Mattick JS, Perera RJ. The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion. Cancer Res. 2011 Jun 1; 71(11):3852-62.

Leucci E, Patella F, Waage J, Holmstrøm K, Lindow M, Porse B, Kauppinen S, Lund AH. MicroRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus. Sci Rep. 2013; 3:2535.

Stein LD. Human genome: end of the beginning. Nature. 2004 Oct 21; 431(7011):915-6.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007 Jun 14; 447(7146):799-816.

Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F, Fan Q. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. 2010 Sep; 38(16):5366-83.

Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010 Aug 12; 466(7308):835-40.

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004 Jan 23;116(2):281-97.

Song L, Han MH, Lesicka J, Fedoroff N. Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. Proc Natl Acad Sci U S A. 2007 Mar 27;104(13):5437-42.

Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, Shiota K. Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem Biophys Res Commun. 2004 Sep 17;322(2):593-600.

Rainier S, Johnson LA, Dobry CJ, Ping AJ, Grundy PE, Feinberg AP. Relaxation of imprinted genes in human cancer. Nature. 1993 Apr 22;362(6422):747-9.

Lottin S, Adriaenssens E, Dupressoir T, Berteaux N, Montpellier C, Coll J, Dugimont T, Curgy JJ. Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. Carcinogenesis. 2002 Nov;23(11):1885-95.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010 Apr 15;464(7291):1071-6.

Chen W, Böcker W, Brosius J, Tiedge H. Expression of neural BC200 RNA in human tumours. J Pathol. 1997 Nov;183(3):345-51.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, et al. FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. Science. 2005 Sep 2;309(5740):1559-63. Erratum in: Science. 2006 Mar 24;311(5768):1713.

Wang P, Yin S, Zhang Z, Xin D, Hu L, Kong X, Hurst LD. Evidence for common short natural trans sense-antisense pairing between transcripts from protein coding genes. Genome Biol. 2008;9(12):R169.

Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009 Feb 20;136(4):629-41.

Matouk IJ, Abbasi I, Hochberg A, Galun E, Dweik H, Akkawi M. Highly upregulated in liver cancer noncoding RNA is overexpressed in hepatic colorectal metastasis. Eur J Gastroenterol Hepatol. 2009 Jun;21(6):688-92.

Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Müller-Tidow C. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene. 2003 Sep 11;22(39):8031-41.

Guffanti A, Iacono M, Pelucchi P, Kim N, Soldà G, Croft LJ, Taft RJ, Rizzi E, Askarian-Amiri M, Bonnal RJ, Callari M, Mignone F, Pesole G, Bertalot G, Bernardi LR, Albertini A, Lee C, Mattick JS, Zucchi I, De Bellis G. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC Genomics. 2009 Apr 20;10:163.

Zhao Y, Guo Q, Chen J, Hu J, Wang S, Sun Y. Role of long non-coding RNA HULC in cell proliferation, apoptosis and tumor metastasis of gastric cancer: a clinical and in vitro investigation. Oncol Rep. 2014 Jan; 31(1):358-64.

Wang W, Lin H, Zhou L, Zhu Q, Gao S, Xie H, Liu Z, Xu Z, Wei J, Huang X, Zheng S. MicroRNA-30a-3p inhibits tumor proliferation, invasiveness and metastasis and is downregulated in hepatocellular carcinoma. Eur J Surg Oncol. 2013 Nov 19. pii: S0748-7983(13)00912-8.

Ma Y, Zhang P, Yang J, Liu Z, Yang Z, Qin H. Candidate microRNA biomarkers in human colorectal cancer: systematic review profiling studies and experimental validation. Int J Cancer. 2012 May 1;130(9):2077-87.